DOCUMENT RESUME

ED 109 685 CS 202 166

AUTHOR O'Donnell, Roy C.
TITLE A Comparison of Two Indices of Syntactic Complexity. Studies in Language Education Report No. 20.
INSTITUTION Georgia Univ., Athens. Dept. of Language Education.
PUB DATE Sep 75
NOTE 13p.

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Comparative Analysis; English Instruction; Evaluation Methods; Grade 9; Language Development; *Language Research; *Language Skills; *Measurement Instruments; Secondary Education; *Syntax; Writing Skills; Written Language
IDENTIFIERS *T Unit

ABSTRACT

The need for an objective index of language development has long been recognized, and various means of measuring language growth have been proposed. This study compares two of the indexes that have recently been devised for measuring language growth. Its specific purposes are (1) to examine the correlation between mean T-unit length and syntactic density scores and (2) to compare mean T-unit length and syntactic density scores as indicators of the grade level of writers. Thirty-two ninth graders were asked to write a composition about the best or worst hour of their day. These compositions were then analyzed for mean T-unit length, mean clause length, and mean number of clauses per T-unit, and were also analyzed for the items needed to compute the syntactic density score. It was concluded that, in spite of its lack of precision, T-unit length is a more useful and usable index of syntactic development than the syntactic density score; and it appears that mean T-unit length is a better indicator of grade level than is the syntactic density score. (RB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED109685

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A COMPARISON OF TWO INDICES OF SYNTACTIC COMPLEXITY

by
Roy C. O'Donnell

Studies in Language Education, Report No. 20
Department of Language Education, The University of Georgia
Athens, Georgia

September 1975

BEST AVAILABLE COPY

CS 202 166

The need for an objective index of language development has long been recognized, and various means of measuring language growth have been proposed (O'Donnell, 1975). It is the purpose of the study reported here to compare two of the indices that have recently been devised.

TWO INDICES OF COMPLEXITY

Mean Length of T-Unit

In a study of grammatical structures written by pupils in grades four, eight, and twelve, Hunt (1965) gave a critical evaluation of some traditional indices of language development and proposed a new index. He pointed out the fact that inadequate punctuation and indiscriminate use of and makes sentence length an unreliable index and proposed the use of a syntactic unit consisting of one main clause and any subordinate clauses attached to the main clause. This unit is grammatically capable of being considered a sentence; and since it is a terminable unit, he proposed that it be called a T-unit. This unit can be identified objectively, and it is not affected by poor punctuation; it has the added advantage of preserving all the subordination achieved by the student and all of his coordination of words, phrases, and subordinate clauses. It does not preserve the student's coordination of main clauses. However, since excessive coordination may be evidence of immaturity rather than maturity, the elimination of coordinated main clauses from the syntactic unit is considered to be a gain rather than a loss.

Hunt called attention to the fact that LaBrant counted coordinated verbs as separate clauses; he contended that only a structure with a subject and

a finite verb should be regarded as a clause. Then, having identified a syntactic unit more useful than the sentence and having limited the definition of the clause, Hunt proposed a revision of the subordination index. He proposed that, instead of dividing the number of subordinate clauses by the total number of clauses as LaBrant had done, the total number of clauses be divided by the number of main clauses. This procedure gives the mean number of clauses per T-unit, which can be converted into the ratio of subordinate clauses to main clauses. For example, a writing sample with an average of 1.5 clauses per T-unit has twice as many main clauses as subordinate clauses.

Hunt computed these indices for the materials analyzed in his study and found statistically significant increases in T-unit length, clause length, and number of clauses per T-unit from grade four to grade eight to grade twelve. He pointed out that these three measures are interrelated, i.e., T-units can be lengthened by either or both of two means: (1) lengthening clauses or (2) increasing the number of subordinate clauses. Thus Hunt's analysis, based on the techniques of transformational-generative grammar, revealed that syntactic complexity of children's language is reflected in both clause length and number of clauses per T-unit. He demonstrated that T-unit length can be computed by multiplying mean words per clause by number of clauses per T-unit; and since T-unit length incorporates the other two indices, it was evident that T-unit length would be useful index of syntactic complexity of language.

O'Donnell, Griffin, and Norris (1967) demonstrated that T-unit length can be used as an index of oral as well as written language complexity. In their study of children's syntax, they computed both mean T-unit length and mean number of sentence-combining transformations per T-unit.

These investigators found a positive relationship between number of words per T-unit and number of sentence-combining transformations per T-unit.

Since the two measures appeared to be of comparable value as indices of syntactic development, they indicated a preference for the more economical of the two and concluded that "mean length of T-units has special claim to consideration as a simple, objective, valid indicator of development in syntactic control" (pp. 98-99).

An obvious limitation of T-unit length lies in the fact that indices based on mean length of syntactic units do not discriminate among the various ways length can be achieved. The T-unit is more satisfactory in this respect than the sentence; but the T-unit can be lengthened in a variety of ways, some of which require much more linguistic maturity than others. The desire to find a more discriminating index of language development has led researchers to attempt to develop more sensitive indices of structural growth.

Syntactic Density Score

One approach to devising a more discriminating linguistic index is reported by Golub and Kidder (1974). Their Syntactic Density Score (SDS) reflects measures of T-unit length, clause length, and number of subordinate clauses; it also takes into account uses of complex verb phrase expansions and various kinds of embedded structures.

Golub and Kidder report that the first step in developing this measure of syntactic density was to determine how children use specific syntactic structures and which ones seem to predict whether their writing will be rated high, medium, or low by their teachers. The items included in the score were selected from sixty-three structures which had been subjected to multivariate analysis. The variable which correlated highly with teacher's ratings of written language samples were selected. Canonical correlation analysis was performed and relative weights were assigned to variables according to their contribution to syntactic density.

The Syntactic Density Score is based on the ten items selected, including words per T-unit, subordinate clauses per T-unit, words per main clause, words per subordinate clause, number of modals, number of be and have forms, number of prepositional phrases, number of possessives, number of adverbs of time, and number of gerunds, participles and unbound modifiers. Each item was assigned a loading ranging from .20 for words per main clause to .95 for words per T-unit. The loading for each item multiplied by frequency results in products which can be summed and divided by the total number of T-units. The quotient is the Syntactic Density Score, which can be converted to a grade level. A score of .5 is equivalent to grade level 1, and a score of 10.9 is equivalent to grade level 14. Each grade level differs from the one below it by a Syntactic Density Score value of .8. Grade level equivalents for Syntactic Density Scores and mean length of T-unit are presented in Table 1.

Table 1
GRADE LEVEL EQUIVALENTS FOR SYNTACTIC DENSITY SCORES AND MEAN LENGTH OF T-UNIT

Grade	Syntactic Density Score	Mean Length of T-Unit
1	.5	--
2	1.3	--
3	2.1	--
4	2.9	8.60
5	3.7	9.32*
6	4.5	10.05*
7	5.3	10.78*
8	6.1	11.50
9	6.9	12.22*
10	7.7	12.95*
11	8.5	13.68*
12	9.3	14.40
13	10.1	--
14	10.9	--

*Interpolated from data obtained by Hunt (1965) for grades 4, 8, and 12

Golub and Kidder indicate that a 500 word language sample can be scored in a half-hour or so; but for convenience they have written a computer program, which they believe has a number of valuable uses:

> The computer program is useful for classroom teachers who want to diagnose entering behavior levels and subsequent performance levels of their students and to match these levels with reading materials used in class. Examination by a teacher of English of the subscores of a class or students gives an indication of strengths and weaknesses that can guide the teacher's planning of objectives and procedures. The Syntactic Density Score of the variety of novels, textbooks and other reading materials used in specially designed mini-courses or performance-based curriculums can be determined quickly, accurately, and inexpensively (p. 1131).

The Syntactic Density Score is capable of discriminating among types of structures. Although it appears to have been developed by empirical procedures, the items included have a high degree of redundancy in what they measure. Words per T-unit results from the combined effects of number and length of clauses; length of clauses, in turn, results in part from the remaining items on the instrument. This redundancy suggests that there might be a high correlation between Syntactic Density Score and words per T-unit.

Although Golub's Syntactic Density Score has some evidence of validity, it seems likely that the measure is less precise than it appears to be, particularly when converted to grade level equivalents. Perhaps in some instances the less expensive measure of mean words per T-unit might be as useful as Syntactic Density Score.

PURPOSES AND PROCEDURES

Purposes

The specific purposes of this study are: (1) to examine the correlation between mean T-unit length and Syntactic Density Scores, and (2) to compare mean T-unit length and Syntactic Density Scores as indicators of grade level of writers.

Procedures

Thirty-two ninth graders were asked to write a composition about the best or worst hour of their day. The thirty-two papers were analyzed for mean T-unit length, mean clause length, and mean number of clauses per T-unit. They were also analyzed for the additional items needed to compute the Syntactic Density Score. The data resulting from the syntactic analysis are presented in Table 2.

Table 2
SELECTED INDICES OF SYNTACTIC COMPLEXITY OF PAPERS
WRITTEN BY THIRTY-TWO NINTH-GRADE PUPILS

Pupil	Total Words	Mean Words Per T-Unit	Mean Words Per Clause	Mean Clauses Per T-Unit	Syntactic Density Score
1	452	22.60	8.69	2.60	4.20
2	308	14.67	9.06	1.62	3.46
3	209	10.45	8.04	1.30	2.12
4	123	7.24	4.73	1.53	2.35
5	222	9.25	7.16	1.29	1.80
6	263	11.43	7.73	1.48	2.27
7	440	12.57	8.80	1.43	2.47
8	284	12.91	9.16	1.41	2.80
9	261	15.35	9.32	1.65	4.07
10	324	10.45	7.90	1.32	2.04
11	230	16.43	8.52	1.93	3.88
12	173	12.36	9.61	1.38	3.23
13	340	13.07	11.33	1.15	2.99
14	247	12.35	9.88	1.25	2.85
15	330	13.20	8.46	1.56	2.80
16	366	11.09	7.32	1.51	2.35
17	357	12.31	8.92	1.38	2.50
18	203	15.61	10.15	1.54	4.34
19	256	12.80	8.26	1.55	2.92
20	182	9.10	7.58	1.20	1.68
21	248	11.81	8.27	1.43	1.87
22	167	11.93	9.28	1.28	2.89
23	177	8.04	6.81	1.18	1.65
24	255	15.94	9.80	1.62	3.88
25	231	14.44	8.88	1.62	2.90
26	203	15.61	10.15	1.54	4.21
27	442	15.78	8.18	1.93	3.01
28	225	14.06	9.00	1.56	3.18
29	245	13.61	7.90	1.72	2.92
30	313	13.04	10.10	1.29	2.95
31	454	14.64	9.66	1.52	2.47
32	181	15.08	10.05	1.50	3.93
Mean	272.22	13.10	8.71	1.51	2.90

The thirty-two papers were rank ordered by mean number of words per T-unit and mean Syntactic Density Score, and a Spearman Rank Order Correlation (Rho) value was computed. The rank ordered distribution is shown in Table 3.

Table 3
RANK ORDERED DISTRIBUTION OF THIRTY-TWO NINTH GRADE PUPILS ON MEAN T-UNIT LENGTH AND SYNTACTIC DENSITY SCORES

Pupil	Rank Order on T-Unit Length	Rank Order on Syntactic Density Score
1	1	3
2	9	8
3	27.5	27
4	32	24.5
5	29	30
6	25	26
7	19	22.5
8	17	19.5
9	7	4
10	27.5	28
11	2	6.5
12	20	9
13	15	12
14	21	18
15	14	19.5
16	26	24.5
17	22	21
18	5.5	1
19	18	14.5
20	30	31
21	24	29
22	23	17
23	31	32
24	3	6.5
25	11	16
26	5.5	2
27	4	11
28	12	10
29	13	14.5
30	16	13
31	10	22.5
32	8	5
	Rho = .88	

Syntactic Density Scores were converted to grade level equivalents by means of the conversion table prepared by Golub and Kidder (1974). Grade level equivalents of T-unit scores were approximated by referring to the

mean scores obtained by Hunt (1965) for pupils in grades four (8.60), eight (11.50), and twelve (14.40). Grade level equivalents for each of the thirty-two pupils on both indices are given in Table 4.

Table 4
GRADE LEVEL EQUIVALENTS OF SYNTACTIC DENSITY SCORES AND MEAN WORDS PER T-UNIT FOR THIRTY-TWO NINTH-GRADE PUPILS

Pupil	Syntactic Density Score	Grade Level Equivalent	Mean Words Per T-Unit	Grade Level Equivalent
1	4.20	5+	22.60	12+
2	3.46	4+	14.67	12+
3	2.12	3+	10.45	6+
4	2.35	3+	7.24	3+
5	1.80	2+	9.25	4+
6	2.27	3+	11.43	7+
7	2.47	3+	12.57	9+
8	2.80	3+	12.91	9+
9	4.07	5+	15.35	12+
10	2.04	2+	10.45	6+
11	3.88	5+	16.43	12+
12	3.23	4+	12.36	9+
13	2.99	4+	13.07	10+
14	2.85	3+	12.35	9+
15	2.80	3+	13.20	10+
16	2.35	3+	11.09	7+
17	2.50	3+	12.31	9+
18	4.34	5+	15.61	12+
19	2.92	4+	12.80	9+
20	1.68	2+	9.10	4+
21	1.87	2+	11.81	8+
22	2.89	3+	11.93	8+
23	1.65	2+	8.04	3+
24	3.88	5+	15.94	12+
25	2.90	4	14.44	12+
26	4.21	5+	15.61	12+
27	3.01	4+	15.78	12+
28	3.18	4+	14.06	11+
29	2.92	4+	13.61	10+
30	2.95	4+	13.04	10+
31	2.47	3+	14.64	12+
32	3.93	5+	15.08	12+
Mean	2.90	4	13.10	10+
Median	2.90	4	12.98	10+
s.d.	.74		2.87	

FINDINGS AND CONCLUSIONS

Findings

One purpose of this study was to examine the correlation between mean T-unit length and Syntactic Density Scores obtained from papers written by thirty-two ninth-grade pupils. The papers were rank ordered on both measures and a Spearman Rho was computed. The resulting Rho of .88, as well as the rank order of each pupil on mean T-unit length and Syntactic Density Score, is shown in Table 3.

A second purpose of the study was to compare mean T-unit length and Syntactic Density Score as indicators of grade level. Data presented in Table 4 show a mean Syntactic Density Score value of 2.90, which is precisely the score equivalent to fourth grade level on Golub and Kidder's conversion table; the mean T-unit score for the entire group is 13.10, which falls about half-way between Hunt's eighth and twelfth grade "norms" and is assumed to be approximately tenth grade level.

Conclusions

Since T-unit length is highly correlated with Syntactic Density Score (Rho = .88) and can be computed for a fraction of the cost of the more complex measure, it is only when information on specific structures is desired that such detailed analysis would be justified. Thus it appears that, in spite of its lack of precision, T-unit length is for many purposes a more useful and useable index of syntactic development than the Syntactic Density Score.

Although no standardized test scores were collected for the ninth graders represented in this study, informally collected evidence indicates that they were not below average in scholastic and linguistic abilities. Therefore, it appears that mean T-unit length is a better indicator of grade level than

is Syntactic Density Score, particularly since no subject scored as high as sixth grade on Syntactic Density Score and only ten scored below ninth grade on mean length of T-unit.

Since the Syntactic Density Score formula gives separate loadings for specific kinds of structures, the resulting scores are likely to be sensitive to different modes of discourse and to different styles of individual writers. For some purposes, such sensitivity would be desirable, but it would seem to lessen the reliability of grade level equivalent scoring. Given the diversity of individual writing styles, the diverse demands imposed on language by varying situations, and various other factors, it is difficult to see how Syntactic Density Score grade equivalents can be used with much confidence.

REFERENCES

Golub, Lester S. and Carole Kidder. 1974. "Syntactic Density and the Computer," Elementary English, 51:8 (November--December), 1128--1131.

Hunt, Kellogg, W. 1965. Grammatical Strucutres Written at Three Grade Levels, NCTE Research Report No. 3. Champaign, Illinois: National Council of Teachers of English.

O'Donnell, Roy C., William J. Griffin, and Raymond C. Norris. 1967. Syntax of Kindergarten and Elementary School Children. NCTE Research Report No. 8, Champaign, Illinois: National Council of Teachers of English.

O'Donnell, Roy C. 1975. A Review of Some Indices of Language Development. Studies in Language Education, Report No. 15, Department of Language Education, The University of Georgia, Athens, Georgia.